

# Entre TBX et Ontolex-Lemon : Quelles Nouvelles Perspectives en Terminologie ?

Silvia Piccini<sup>1</sup>, Federica Vezzani<sup>2</sup> et Andrea Bellandi<sup>1</sup>

<sup>1</sup> Istituto di Linguistica Computazionale “A. Zampolli”, Area della Ricerca CNR di Pisa, Via Giuseppe Moruzzi, 1, 56124 Pisa PI, Italia

<sup>2</sup> Università degli Studi di Padova, Dipartimento di Studi Linguistici e letterari, Via Elisabetta Vendramini, 13 35137 Padova PD, Italia

## Abstract

Cet article porte sur une analyse contrastive multi-niveaux des technologies à la base de TBX et de *Ontolex-lemon* afin de modéliser les données terminologiques multilingues au sein de ressources terminologiques.

## Keywords

Terminologie, TermBase eXchange, Ontolex-Lemon

## 1. Résumé étendu

Le développement de ressources terminologiques et leur maintenance optimale sont des activités laborieuses et chronophages qui nécessitent souvent d’investissements économiques substantiels. C’est la raison pour laquelle l’élaboration de normes pour la représentation et l’échange de données occupent une place essentielle dans le domaine de la terminologie. L’interopérabilité<sup>2</sup> entre les formats terminologiques est assurée par l’adoption de la norme ISO 30042 : 2019 qui promeut l’adoption du format *TermBase eXchange* (TBX), à savoir une famille XML de formats d’échange terminologique basée sur le méta-modèle structurel hiérarchique *Terminological Markup Framework* (TMF) (ISO-16642 : 2017).

Au cours des dernières années, des solutions intéressantes ont été proposées afin de favoriser l’interopérabilité par le biais de technologies du Web Sémantique (WS) [2] et des principes promus par l’initiative des « données liées » (*Linked data*) [3, 5], qui permettent de partager et de réutiliser les données entre différentes applications et de les relier aux ressources disponibles sur le Web. Dans cette optique, le modèle *Ontolex-Lemon* (<https://www.w3.org/2016/05/ontolex/>) a été développé en s’imposant en tant que standard *de facto* pour la représentation des ressources linguistiques et terminologiques dans le WS. Les avantages offerts par ces technologies sont incontestables. Premièrement, elles assurent la trouvabilité, l’accessibilité, l’interopérabilité et la réutilisabilité des données de la recherche (de l’acronyme anglais FAIR) [31] conformément à la philosophie de la science ouverte et aux stratégies de recherche et d’innovation de la Commission européenne pour la période 2020-2024. Deuxièmement, le modèle de données RDF, à la base du WS, permet d’atteindre non seulement une interopérabilité syntaxique – déjà garantie par la structure de données XML à la base de TBX – mais également une interopérabilité sémantique, assurant ainsi un échange de données caractérisées par une sémantique désambiguïsée et partagée.

L’utilisation des technologies du WS signifie en effet représenter des données terminologiques sous forme de triplets RDF et structurer formellement la sémantique de ces données en ontologies au moyen

---

*1st International Conference on “Multilingual digital terminology today. Design, representation formats and management systems”, June 16 – 17, Padova, Italy*

EMAIL: [silvia.piccini@ilc.cnr.it](mailto:silvia.piccini@ilc.cnr.it) (A. 1); [federica.vezzani@unipd.it](mailto:federica.vezzani@unipd.it) (A. 2); [andrea.bellandi@ilc.cnr.it](mailto:andrea.bellandi@ilc.cnr.it) (A. 3)

ORCID: 0000-0002-2584-0191 (A. 1); 0000-0003-2240-6127 (A. 2); 0000-0002-1900-5616 (A. 3)



© 2022 Copyright for this paper by its authors.  
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

<sup>2</sup> Cf. [30].

du langage *Web Ontology Language* (OWL). Ce dernier, basé sur un sous-ensemble de la logique du premier ordre, rend les machines capables de comprendre et d'interpréter les données et d'effectuer des inférences logiques, déduisant ainsi de nouveaux faits. Ainsi, de nombreuses solutions ont été proposées dans la littérature afin de transformer une structure de données XML en une formalisation ontologique plus expressive [1, 4, 6] permettant, de cette façon, de définir les relations hiérarchiques qui existent entre deux ou plusieurs concepts, de préciser les propriétés des relations (transitivité, symétrie, réflexivité, etc.) et les contraintes (ou axiomes) générales et formelles entre les concepts, les attributs et les relations, et enfin d'élaborer des règles permettant de déduire de nouvelles connaissances [20]. Il n'est donc pas surprenant que la terminologie ait commencé à s'intéresser à ces technologies, comme en témoigne le nombre croissant de jeux de données ouverts et interconnectés publiés dans le cadre du Web de données [26, 27] ou les études et les outils visant à transformer et à publier des terminologies en tant que « données liées » [10, 13]. À cet égard, sont indicatives les solutions engagées en appui à la Stratégie européenne d'interopérabilité (EIS) visant à transformer en RDF deux ressources d'intérêt stratégique pour la Communauté européenne telles que IATE [29] et EuroVoc [11].

Les avantages technologiques incontestés ont souvent constitué la force motrice vers le paradigme des données liées comme en témoigne la croissance constante du soi-disant *Linguistic Linked Open Data Cloud* (LLOD) [8]. Cependant, les avantages ne se limitent pas seulement à ces aspects. Cet article vise à étudier les effets théoriques que l'adoption des technologies du WS et du paradigme des « données liées » peut entraîner dans les études terminologiques en tant que pistes de réflexion qui, à notre avis, n'ont pas encore été suffisamment explorées. Le format TBX reflète, en effet, la vision traditionnelle de la terminologie qui remonte aux travaux d'Eugen Wüster [32, 33].

Bien qu'une description précise et détaillée des aspects linguistiques d'un terme puisse être accomplie au moyen de catégories de données spécifiques (ISO 12620 : 2019), le concept est généralement vu comme une entité simple et universellement partagée selon une vision de la réminiscence aristotélicienne (τὰ πάντα πᾶσι). Cette perspective, qui est à la base des principales bases de données terminologiques mises en place autour des années 70/80 (i.e. EURODICAUTOM, IATE, TERMIUM Plus, etc.), est très efficace pour les tâches de traduction. Par conséquent, il n'est pas surprenant que les outils de TAO soient capables d'exporter et/ou d'importer des glossaires dans le format TBX. Cependant, comme l'ont montré de nombreuses approches terminologiques [12, 7, 19, 16, 28, 24], le concept doit plutôt être vu comme une entité complexe et multiforme, à savoir une liste des conditions nécessaires et suffisantes résultant de catégorisations qui peuvent varier intra- et inter-linguistiquement ainsi que diachroniquement.

Comme les études variationnistes l'ont mis en évidence [17, 18, 15, 14] les langues peuvent se focaliser sur le concept sous différents angles : certains traits peuvent se traduire en sens dans une langue mais rester dans l'ombre dans une autre langue (tout en restant accessible à la pensée et à l'expérience), donnant lieu à des phénomènes d'anisomorphisme. La variation peut également être intra-linguistique : au sein d'une même langue de spécialité, un concept peut se matérialiser en plusieurs dénominations synonymes et dans chacune d'elles les caractéristiques du concept considérées comme cognitivement plus saillantes se cristallisent et se transforment en sens lexical.

Comme nous essaierons de le (de)montrer, l'utilisation des technologies du WS, telles que les ontologies, devient fondamentale pour décrire la complexité à la fois du concept et du sens d'un terme et ainsi faire comprendre aux chercheurs la dynamique qui se cache derrière les phénomènes de la polysémie, de la synonymie, de l'anisomorphisme et de la variation diachronique.

En d'autres termes, les technologies du WS sont particulièrement efficaces pour capturer et décrire les concepts dits endocentriques [25]. Ces derniers † à la différence des concepts exocentriques ancrés dans une expérience partagée par des individus au-delà des frontières linguistiques particulières † dépendent fortement du système de relations lexicales propre à une langue, et ne peuvent être pleinement compris sans une description qui prend en compte le caractère complexe des concepts et la dimension du sens du terme.

Après une analyse contrastive multi-niveaux (syntaxe, sémantique, tâches, outils) des technologies à la base de TBX et de *Ontolex-lemon*, des études de cas extraites à partir du domaine médical seront illustrées. En particulier, nous examinerons des cas de variation diachronique de conceptualisation (par exemple pour le concept d'« hystérie »), de variation socio-culturelle (le cas du terme « SIDA ») et de variation interlinguistique liée à des aspects connotatifs (français « sein »/« mamelle » ~ italien *seno/mammella*) qui peuvent influencer la perception des termes/concepts en perspective multilingue.

Enfin, pour favoriser une synergie profonde entre les ressources en TBX et les ressources en RDF, nous proposerons un outil pour convertir automatiquement les ressources créées dans la version la plus récente de TBX (de 2019) en RDF (et notamment en *Ontolex-Lemon*). À l'heure actuelle, le seul outil existant TBXtoRDF créé par Cimiano et al. (2015) n'est compatible qu'avec une version désormais obsolète de TBX (2008) et de *Ontolex-Lemon*. En ce sens, nous proposons donc un système qui prend en charge toutes les spécificités de la nouvelle version de TBX datée de 2019 et les dernières spécifications de *Ontolex-Lemon*.

## 2. Références

- [1] R. R. Amorim, M. Lama, E. Sánchez, A. Riera, X. A. Vila, A learning design ontology based on the IMS specification, *Journal of Educational Technology & Society* 9(1) (2006) 38-57.
- [2] T. Berners-Lee, J. Hendler, O. Lassila, The semantic web, *Scientific American* 284(5) (2001) 34-43.
- [3] T. Berners-Lee, Linked data-design issues, 2006. URL: <http://www.w3.org/DesignIssues/LinkedData.html>.
- [4] S. Bischof, N. Lopes, A. Polleres, Improve Efficiency of Mapping Data between XML and RDF with XSPARQL, in: *Proceedings of Web Reasoning and Rule Systems – Fifth International Conference, RR2011*, volume 6902 of *Lecture Notes in Computer Science (LNCS)*, Springer-Verlag, Galway, Ireland, August, 2011, pp. 232–237.
- [5] Ch. Bizer, T. Heath, T. Berners-Lee, Linked data: The story so far, in: A. Sheth (Ed.), *Semantic services, interoperability and web applications: emerging concepts*, 1st ed., IGI Global, Hershey, Pennsylvania, 2011, pp. 205-227.
- [6] H. Bohring, S. Auer, Mapping XML to OWL Ontologies, in: *Proceedings of Marktplatz Internet: Von e-Learning bis e-Payment*, 13. Leipziger Informatik-Tage (LIT 2005), Gesellschaft für Informatik e. V., Bonn, 2005, pp. 147-156.
- [7] M.T. Cabré-Castellvi, Theories of Terminology. Their description, prescription and explanation, *Terminology* 9(2), (2003) 163-200.
- [8] Ch. Chiarcos, S. Nordhoff, S. Hellmann, *Linked Data in Linguistics*, Springer, Heidelberg, 2012.
- [9] Ph. Cimiano, Ch. Chiarcos, J. McCrae, J. Gracia, *Linguistic Linked Data*, Springer International Publishing, 2000.
- [10] Ph. Cimiano, et al., Linked terminologies: applying linked data principles to terminological resources, in: *Proceedings of the eLex 2015 Conference*, 2015.
- [11] D. Dechandon, A. Gerencsér, M. R. Ruiz, Terminology: Towards a Systematic Integration of Semantics and Metadata, in: *Translating and the Computer 41*, *Proceedings of AsLing (The International Association for Advancement in Language Technology)*, Editions Tradulex, Geneva, 2020.
- [12] M. Diki-Kidiri, Une approche culturelle de la terminologie, *Terminologies nouvelles* 21 (2000) 27-41.
- [13] M.P. Di Buono, et al., Terme-a-llod: Simplifying the conversion and hosting of terminological resources as linked data, in: *Proceedings of the 7th Workshop on Linked Data in Linguistics (LDL-2020)*, 2020.
- [14] P. Drouin, A. Francœur, J. Humbley, A. Picton, Multiple Perspectives on Terminological Variation, *Terminology and Lexicography Research and Practice*, 18, John Benjamins, Amsterdam, 2017.
- [15] P. Dury, Que montre l'étude de la variation d'une terminologie dans le temps. Quelques pistes de réflexion appliquées au domaine médical, *Debate Terminológico* 9 (2013) 2-10.
- [16] P. Faber, A cognitive linguistics view of terminology and specialized language, *Series: Applications of Cognitive Linguistics [ACL]* 20, 1st ed., De Gruyter Mouton, Berlin, Boston, 2012.
- [17] J. Freixa, Causes of denominative variation in terminology: A typology proposal, *Terminology* 12(1) (2006) 51-77.

- [18] J. Freixa Aymerich, S. Fernández Silva, M. T. Cabré Castellví, La multiplicité des chemins dénommatifs, *Meta*, LIII(4) (2008) 730-747.
- [19] F. Gaudin, *Socioterminologie. Une approche sociolinguistique de la terminologie*, Duculot, Bruxelles, 2003.
- [20] Y. Gil, V. Ratnakar, A Comparison of (Semantic) Markup Languages, in: *Proceedings of the Fifteenth International FLAIRS Conference*, Pensacola Beach, AAAI Press, Florida, 2002, pp. 413-418.
- [21] ISO-12620, *Management of terminology resources – Data category specifications*. Standard, International Organization for Standardization, Geneva, CH, 2019.
- [22] ISO-30042, *Management of terminology resources – TermBase eXchange (TBX)*. Standard, International Organization for Standardization, Geneva, CH, 2019.
- [23] ISO-16642, *Computer applications in terminology – Terminological markup framework*. Standard, International Organization for Standardization, Geneva, CH, 2017.
- [24] M. C. L’Homme, *Lexical semantics for Terminology. An Introduction*, John Benjamins Publishing Company, Amsterdam/Philadelphia, 2020.
- [25] M. Prandi, *The building blocks of meaning: ideas for a philosophical grammar*, John Benjamins Company, Amsterdam/ Philadelphia, 2004.
- [26] M. Salvadores, et al., Biportal as a dataset of linked biomedical ontologies and terminologies in RDF, *Semantic Web Journal* 4(3) (2013) 277-284.
- [27] B. Smith, et al., The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration, *Nature biotechnology* 25(11) (2007) 12-51.
- [28] R. Temmermann, *Towards New Ways of Terminology Description: The Sociocognitive-Approach*, John Benjamins Publishing, Amsterdam/Philadelphia, 2000.
- [29] D. Vellutino, et al., Verso l’interoperabilità semantica di IATE. Studio preliminare sul lessico dei Fondi strutturali e d’Investimento Europei (Fondi SIE), *Diversité et identité culturelle en Europe XIII* (1) (2016) 187-204.
- [30] P. Wegner, Interoperability, *ACM Computing Surveys (CSUR)* 28(1) (1996) 285-287.
- [31] M. D. Wilkinson, et al., The FAIR Guiding Principles for scientific data management and stewardship, *Scientific data* 3.1 (2016) 1-9.
- [32] E. Wüster, *The Machine Tool. An Interlingual dictionary of basic concepts*, 1 ed., Technical Press, London, 1968.
- [33] E. Wüster, *Einführung in die allgemeine Terminologielehre und terminologische Lexikographie*, Springer, Wien, 1979.